Coursera Capstone

Recommendation Place To Build a New Hotel in Yogyakarta, Indonesia

By:
Febiana Nur Ramdani
October 27, 2019

Introduction

Every person in this world would want to go on vacation especially when holidays come, some of the people spend their time to get together with their family whether just sit in their house or going somewhere with their family. Yogyakarta is one of the famous cities in Indonesia which is located in Central Java and often visited by many visitors both local visitors or foreign visitors, during holidays the hotel occupancy in Yogyakarta can boost to significant number, based on This article, the Association of Indonesian Hotels and Restaurants (PHRI) in the Special Region of Yogyakarta estimates that occupancy rates of star hotels in this area will reach 90 percent during Eid 2019. If we think in business, this may become a good profit to build a new hotel at strategic places in Yogyakarta, the new hotel built by a company may become huge loss for their business moreover, there are lot of hotels in Yogyakarta with various star they had. However, picking a strategic location with small number of rivals and good location may produce huge profit instead of producing huge loss.

When thinking about a strategic place we need to considering some factors to make the hotel are well placed to catch visitors, the 2 main factors are including:

- **Rival** considering the number of hotel already place in a neighborhood is important because we don't want to build a hotel with lot of another hotel there
- **Location** location are crucial factor to be considered because a location with low number of venue may make the visitors are not interested for staying in our hotel because there will be not many visitor go to that place

Business Problem

As the number of Yogyakarta popularity increases, the number of hotel in Yogyakarta are also increases, a strategy are needed to make the hotel are well placed in a strategic location to prevent the risk of loss. Running hotel business is like gabling with huge bet because the fund to build a hotel especially 5-star hotel with the staff and marketing are not like building a mini market in Yogyakarta, we cannot put a high risk to something without planning and strategy. The goal for this project is to select best location to build a hotel and give an insight about the hotel distribution by clustering the hotel in Yogyakarta using K-means and mapping the clustered hotel and number of hotel that already built in a neighborhood in Yogyakarta using folium map therefore we can choose where is the best place to build the hotel.

Target Audience

The target audience for this project are the investor or company to spread their company wings or build a new hotel in Yogyakarta by considering that Yogyakarta are one of

famous country in Indonesia and visited by many local or foreign visitor to spend their holiday or just come to Yogyakarta in another purpose

Data

The data needed to accomplsh the goal for this project are dataset that obtained from scrapping a website to extract the list of neoghborhood in Yogyakarta and it' s zip code, latitude and longitude of neighborhood in Yogyakarta are also needed to visualize the neighborhood in folium map, however, to get latitude and longitude i need to use a geocoder library in python by inputed the neighborhood name in Yogyakarta in the following format:

Neighbor name, Yogyakarta City, Yogyakarta

Data Source

The problem for this project are there are no dataset of list of neighborhood in yogyakarta city document therefore i need to scrape a website and covert it into a dataframe. However the website are not containing latitude and longitude for each neighborhood and to obtain that i will using geocoder library and add the output into our dataframe. To obtain hotel list in a nighborhood, i will using foursquare API by inputting latitude and longitude to the url, however, an ID and secret ID are needed to act as a valid user of foursquare because foursquare have a imit request to access their datas. You can see the website source Here

Methodology

This project will using datas obtained by scrapping a webpage using beautifulsoup because there are no document such as csv files uploaded to internet, the scrapping are done to extract a list of neighborhood in Yogyakarta inclusing the zip code, below are some of the data obtained from scrapping a website:

	Neighborhood	Postal Code
0	Bausasran	55211
1	Tegal Panggung	55212
2	Suryatmajan	55213
3	Sosromenduran	55271
4	Pringgokusuman	55272

Figure 1. List of neighborhood in Yogyakarta and its zip code

Latitude and longitude are needed to visualize the data into folium map, however, our scrapped website are not containing the latitude and longitude for each neighborhood in Yogyakarta, i need to obtain in using geocoder library and inputting the zip code of each neighborhood and add it into the dataframe, below are the code for getting the latitude and longitude.

```
def get_latlng():
    lat_lng_coords = None

    for post in df['Neighborhood']:
        g = geocoder.arcgis('%s, Yogyakarta City, Yogyakarta' % (post))
        lat_lng_coords = g.latlng
        lat.append(lat_lng_coords[0])
        lng.append(lat_lng_coords[1])
```

Figure 2. Code for getting Latitude and Longitude

After having the latitude and longitude the next task are to retreive the venue near the neighborhood using the foursquare API by inputting the latitude and longitude for each neighborhood, before continuing further process, I need to make sure that the venue category obtained from froursquare API are contning hotel, therefore, I print the unique value from the data of venue category and find if there hotel category, if there then i doing one hot encoding to convert categorical value into binary value because k-means cannot deal with categorical value, then i group the data take the dataframe based on the neghborhood and get the data of the neighborhood and the hotel value only to be processed in clusterng because all I want to do is visualizing the hotels built in neighborhood in Yogyakarta, below are the dataset used for clustering

	Neighborhood	Hotel
0	Baciro	0.038462
1	Bausasran	0.040000
2	Bener	0.020408
3	Brontokusuman	0.138298
4	Bumijo	0.190000

Figure 3. List of data used in clustering

Now the data are ready to be used in clustering algorithm, k-means are using k cluster to define to how many cluster the data will be divided based on their similarity, and to find the optimum k I will used metrics from sklearn called silhouette score

because comparing to elbow method, silhouette method are more suitable to the dataset because it measures how smilar a point is to its own cluster compared to other custer, below are the results of my silhouette score

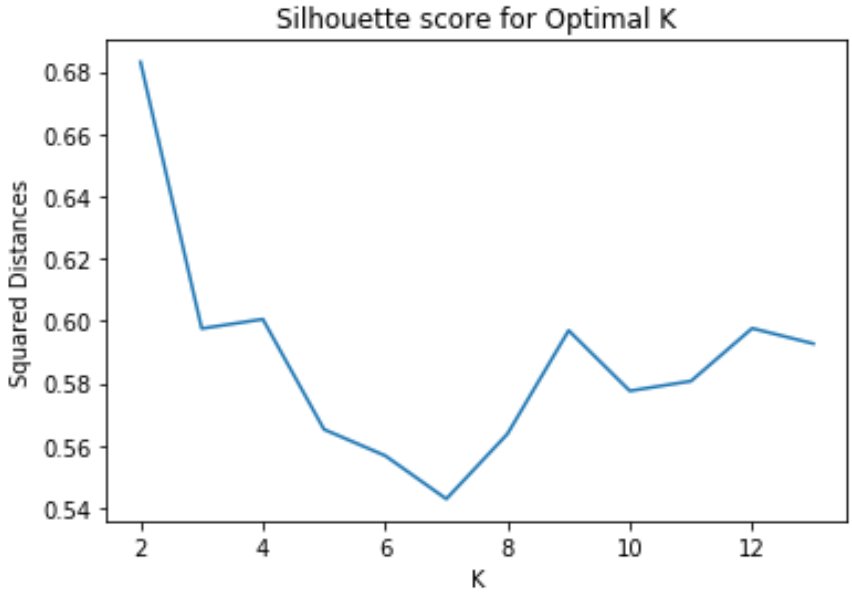

Figure 4. Silhouette score

The figure shows that the optimum k are in 2 clusters, however, 2 cllusters are too small for our choice and we cannot find the best places for the places where new hotel will be build and on the other hand k = 9 are too much and makes us harder to find the best place, therefore, I choose k = 4 because it is the most fitted value for my dataset. After executing clustering using k-means with k value equal to 4, I put the label into the dataframe to be used fot data visualzation later, after that I created a copy of my dataset and join all the column to it then the dataframe will have column: Neighboorhood, Hotel, lat, lng, and Hotel Count

Result

Last part of this project are to visualize the clustered data into folium map and count sum of hotel in each sluster to give me an insight about which cluster that having more hotel compared with another cluster, below are the result:

	Cluster	Hotel Count
0	Cluster 0	96.0
1	Cluster 1	74.0
3	Cluster 3	68.0
2	Cluster 2	32.0

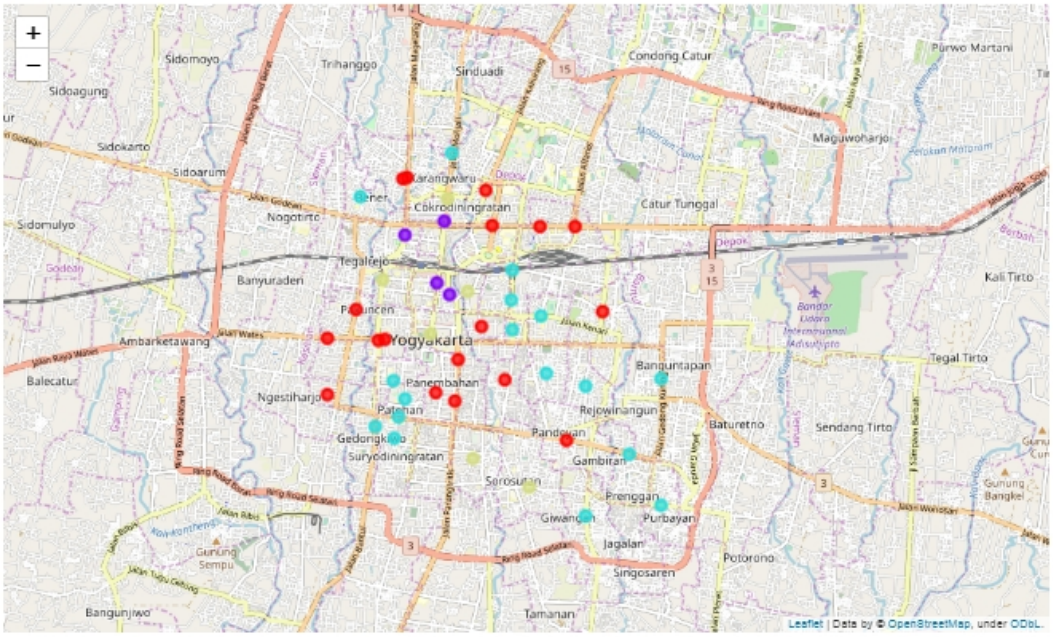

Figure 5. Folium map result

- Cluster 0: Red (96 Hotels)
- Cluster 1: Purple (74 Hotels)
- Cluster 2: Blue (32 Hotels)
- Cluster 3: Yellow (68 Hotels)

Discussion

From the data above, we know that cluster 0 (Red) are having biggest number of hotel built, and on the other side, neighborhood within cluster 2 (Green) are having lesser hotel built, this is happening because the hotel in cluster 1 are placed in the strategic places like near famous spot, have easy access to famous spot or because the area are near to a university and yogyakarta are known as student city because there are lot of university and yogyakarta and the student are come from various districts or even different island, therefore, it can be a potential place to build a hotel. However, i have recommendation to build a hotel, in semaki (cluster 2) there are only 2 hotels built and the area are near to Universitas Gajah Mada and Mandala Krida Stadion which may become potential place to build new hotel

Conclusion

This is the last section of my report and I will conclude this report, this project starts by answering the business problem which is: Where is the recommended place to build new hotel in Yogyakarta when there are lot of hotel already built in that city, the fisrt step to answer that is to collecting the data needed including: Neighborhood name, latitude and longitude of the neighborhood, and the total number of hotel already built in a neighborhood. Then, the data are clustered to k-means and visualize it in folium map to get insight about the result.